Improved Search Techniques

Caleb Ronald Albornoz

NASA Kennedy Space Center

Major: Computer Science

USRP Fall 2012

Date: 7 DEC 2012

Improved Search Techniques

Abstract

Thousands of millions of documents are stored and updated daily in the World Wide Web. Most of the information is not efficiently organized to build knowledge from the stored data. Nowadays, search engines are mainly used by users who rely on their skills to look for the information needed. This paper presents different techniques search engine users can apply in Google Search to improve the relevancy of search results.

1 Introduction

According to the Pew Research Center, the average person spends eight hours a month searching for the right information. For instance, a company that employs 1000 employees wastes $2.5 million dollars on looking for nonexistent and/or not found information. The cost is very high because decisions are made based on the information that is readily available to use. Whenever the information necessary to formulate an argument is not available or found, poor decisions may be made and mistakes will be more likely to occur. Also, the survey indicates that only 56% of Google users feel confident with their current search skills. Moreover, just 76% of the information that is available on the Internet is accurate.

2 NASA – Lessons Learned Search Network

Engineers starting new projects at NASA are required to look for lessons learned on past projects. This process is necessary to understand the successes and failures from previous projects. There is a need to maximize the potential of the Lessons Learned Engine to cover a bigger scope of information relevant for design, development, execution and control groups. Creating a smarter, more efficient and user friendly, and more reliable engine, managers will feel more confident information they share will be efficiently utilized for future applications. To improve the actual capabilities of NASA LLIS, it is necessary to look at its actual potential.

2.1 Current Capabilities

The Lessons Learned application allows NASA engineers to upload reports or lessons from completed projects. After the lesson has been uploaded to the system, the lesson is reviewed for accuracy. Managers from different NASA agencies are required by policy to submit lessons after each project. NASA LLIS can be found at the following address: *https://nen.nasa.gov/web/*nen.Whenever a user looks for lessons, the user is limited to searching with the existing tools.

2.1.1 Search Engine

When performing a search, a query will be searched in the Engineering Repositories. Engineering Repositories is the database that contains all the lessons learned in the system. The platform allows changes in the search domain. Only authorized users are able to search in the NEN site. If you cannot find information in the previous two locations, it is suggested to use NASA Domain which will look for the entry in *htttp://www.nasa.gov*. Additionally, you can be redirected to Google search engine for a more general search. On the other hand, LLIS lets you use the employee locator tool.

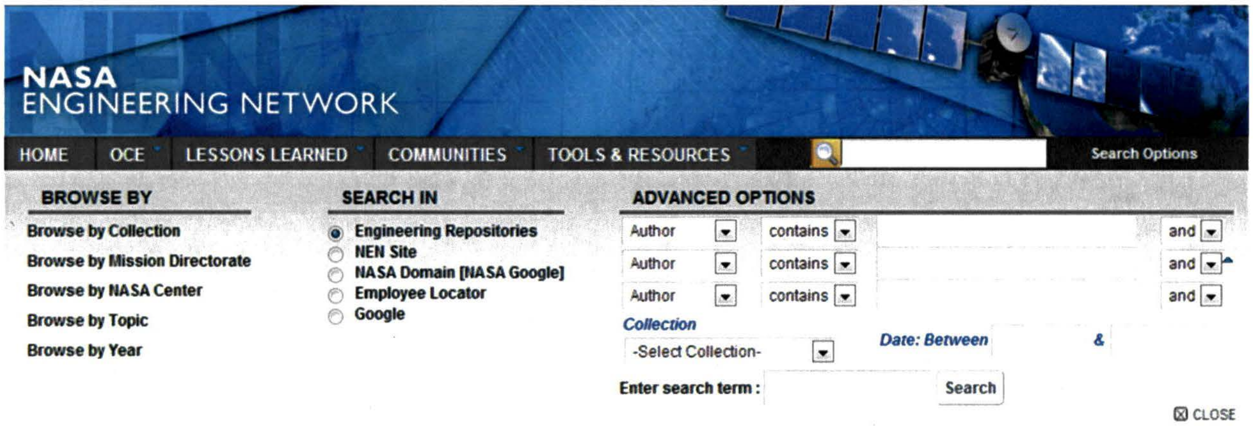


Figure 1. LLIS platform

2.1.2 Multimedia

There are currently three video lessons that can be accessed through the multimedia application. Having more media file tools as available resources to learn may boost the popularity of the LLIS network. Multimedia is a source that let engineers and project managers record or display public presentations from the projects made. It is also considered as a new source to reach out to younger generations.

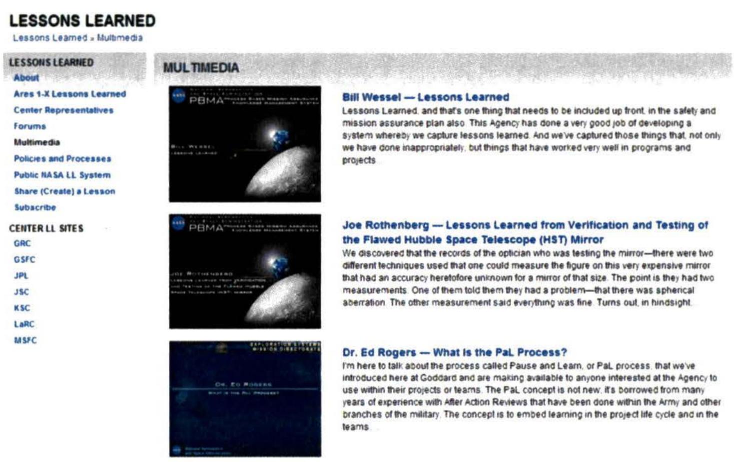
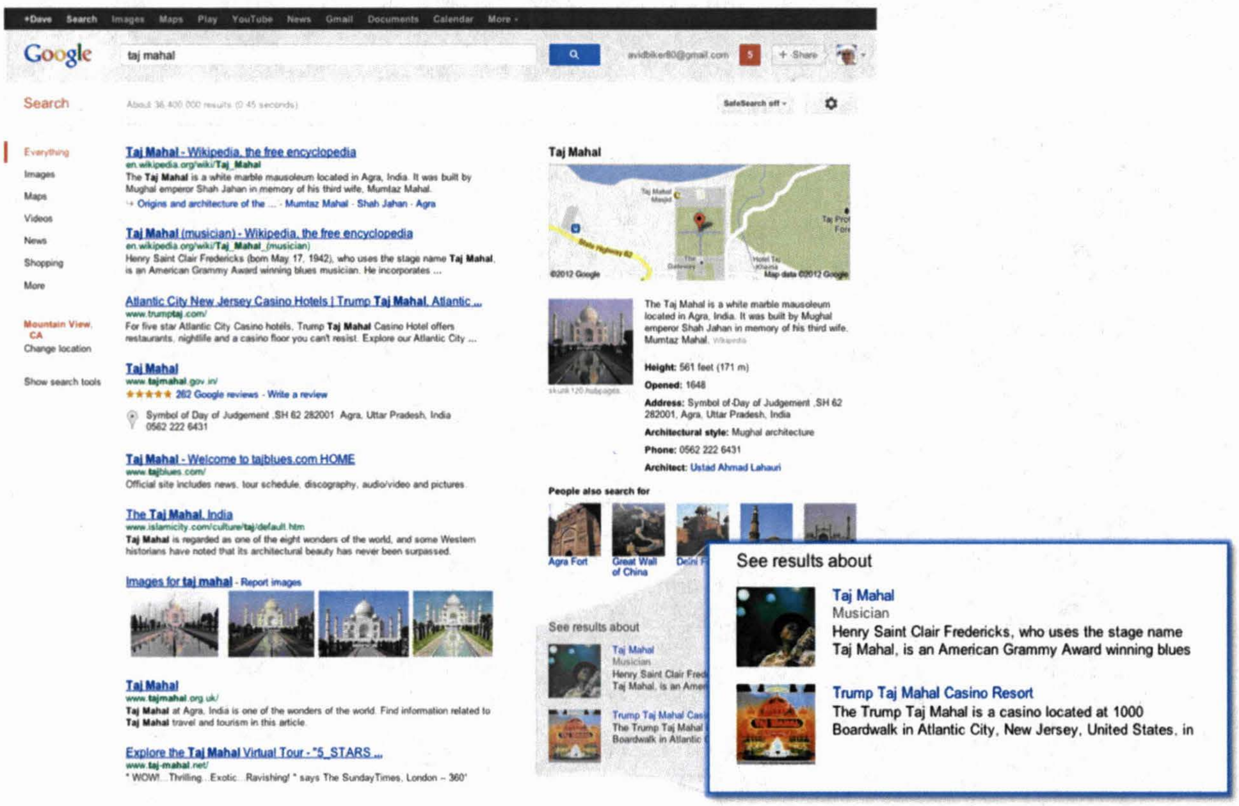


Figure 2. LLIS multimedia environment

3. Google Knowledge-Graph

The World Wide Web has been populated by over a billion websites on the internet. There is information overload. Consequently, there are enough resources to perform smarter strategies to satisfy consumers' needs and build more organized structures. Even though we can only deal with syntax on the web, search engines can support automated reasoning techniques. In May of 2012, Google introduced Knowledge Graph which is a tool built upon ontologies and taxonomies for the semantic web. Knowledge Graph will serve as a model and concept for an advanced and improved version of NASA LLIS.

Google Knowledge graph focuses on three main aspects: find the right thing, get the best summary and go deeper and broader. Knowledge Graph will provide efficiencies when searching for information. Without automated processing and reasoning we cannot obtain useful answers.



4. Semantic Web Development

The Semantic web is an extension of the boundaries that exist in many web pages and applications. The semantic web goes beyond adding meaning or knowledge to a collection of unorganized information. Adding structure to all the information we need will return an answer engine rather than a search one. The graphic below describes how this concept can be accomplished.

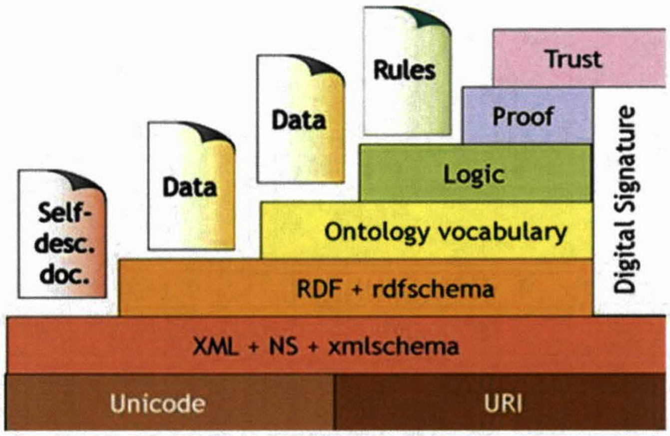


Figure 3. Building semantic knowledge

It is important to consider the structure of the pyramid above. The foundation or base is URI which encompasses Unicode. Unicode is the representation of text expressed in most of the world's writing systems. The URI (Uniform Resource Identifier) bar representation is used to identify a resource. URI (Uniform Resource Identifier) is composed of URL (Uniform Resource Locator) and URN (Uniform Resource Name) which will give us its address/location on the web and its name respectively. After we identify our source of information, we need to encode that information in a format that is human readable and machine readable. To accomplish that goal, the XML (Extensible Markup Language) format is used to define the set of rules. To represent our ontology or knowledge, there is a need to create a metadata data model. RDF (Resource Description Framework) schema is a set of classes used to describe knowledge representation language. It also provides basic elements for the description of ontologies intended to build structures.

The core of ontology is taxonomy. Prior to building ontologies, the information used needs to be classified and grouped in a certain matter. Mainly, taxonomies are structure in the following format: *every A is B*, B represents the more general term. Ontologies use OWL (Web Ontology Language) as a format to build knowledge. OWL formats create triples: subject, predicate and object to analyze syntax and add intelligence to the taxonomy. Logic will analyze the use of valid reasoning for inference and demonstration. The most important step is to link the different ontologies together to create more concrete knowledge via proof. Trust will be the final result of computing databases and information stored in the internet. The ultimate goal is to create an engine that users trust in its efficiency and accuracy.

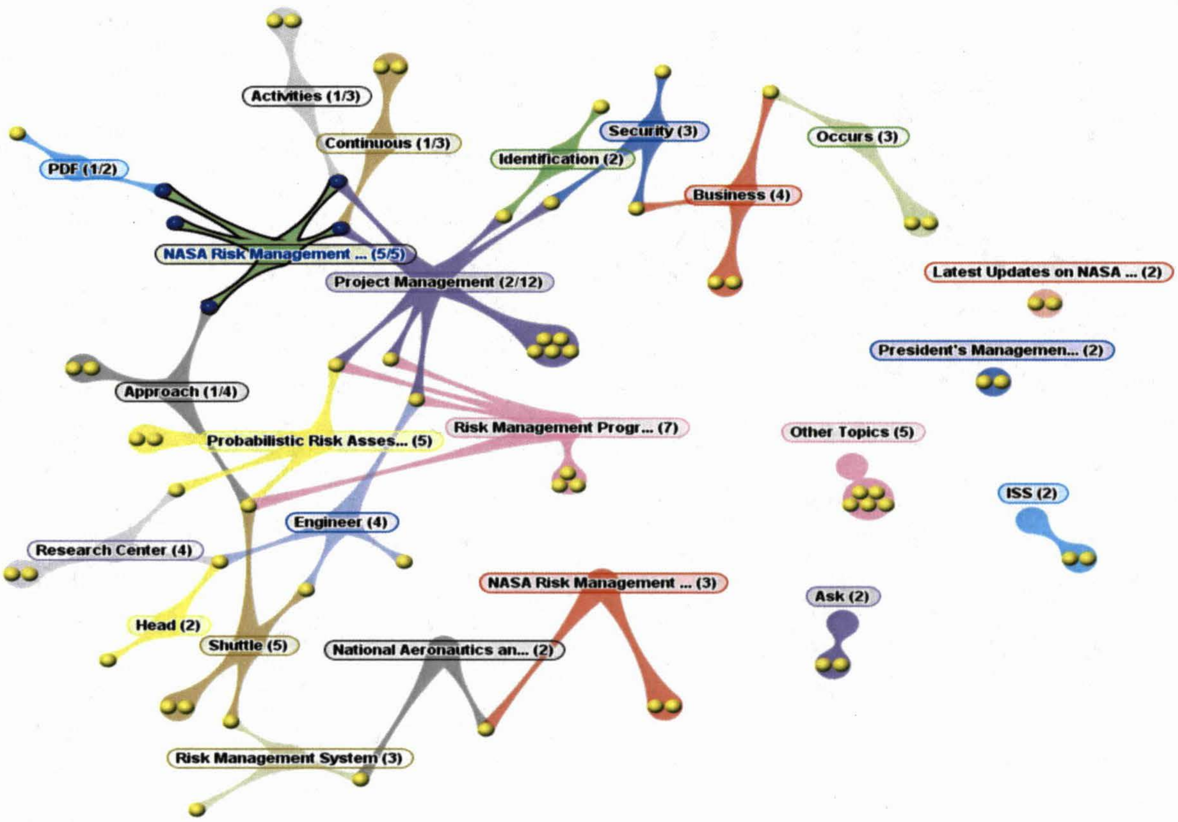


Figure 4. Risk management ontology

5. Knowledge Graph of LLIS

A website has been created to test the application of semantic technologies to the actual LLIS search engine. Creating a more organized environment with Google Custom Search allows the team to differentiate the results produced. The testing site has been uploaded to http://basicse.biz.ly/.

6. Optimizing Search

The strategy utilized was the use of shortcuts to maximize the efficiency of NASA LLIS and Google Search. The listed shortcuts will work in both environments.

- Boolean Operators: they are mostly used when there are more than two keywords to be entered. The AND operator will look for results that contain both keywords on the same result. On the other hand, the OR operator will rather look for either only a keyword appears in one result.

  Syntax, *keyword1* AND *keyword2*, *keyword1* OR *keyword2* (|)

- Specify File Type: Users get the option to narrow their search by a specific type of file. These files can be .xls, .doc, .pdf, .ppt, .html, among others.

  Syntax, *Keyword1 keyword2* filetype:.format

- Exclude: The exclude works as a negation for keywords. For instance, you may be looking for results that contain space missions that do not refer to the Soyuz vehicle. This step filters information according to your needs.

  Syntax, *Keyword1 –keyword2* (exclude "not")

- Quotations Marks: They will be used whenever you want your results to perfectly match the order and words specified in your query.

  Syntax, "*word in keyword with a*" (to specify exact phrase/word)

- Date: Search engines let us filter results for a specific time frame. You may find it useful whenever you look for information between a given period of time.

  Syntax, *keyword1 keyword2 "year1...year2"*

7. Conclusions

Users can use the capabilities that LLIS has to offer. One of them is the advanced search results. If lessons are not submitted, the search engine may not return results for a lesson that does not exist. However, information regarding that topic can be found when selecting "NASA Domain" in the LLIS advanced search box. The LLIS engine will ignore parts of speech words such a conjunctions, prepositions, and articles such as he, she or it.

People can increase the number of relevant resources by making use of the advanced search options, domain choice options and submitting more lessons from their own experiences. After users become more engaged with the benefits of submitting more lessons, search engines will provide with more useful results to future keywords.

When writing a query, the order of words does not matter. Yet, you can make use of quotations marks to truncate the order. Consequently, the detailed suggestions can be found in an acronym format: LLIS (1. List all possible keywords, 2. Leave out parts of speech, 3. Include necessary shortcuts, 4. Specify final terminology) Creating content using smart techniques will add relevance and usefulness to the quality of the NASA Lessons Learned application.

8. References

[1] Spaceport News. Kennedy celebrates 50 years of success. Vol. 52, No. 13. http://www.nasa.gov/centers/kennedy/pdf/664283main_jun29-2012.pdf

[2] Lessons Learned Across the Agency. http://llis.nasa.gov/llis/search/home.jsp

[3] KDP-P-2713, Technical Review Process.

[4] Eric J. Schafer, Gena Henderson, and Damon Stambolian. System Engineering Processes at Kennedy Space Center for Development of SLS and Orion Launch Systems. 2013 IEEEAC paper #2686

[5] Kristen Purcell, Joanna Brenner, Lee Rainie. Search Engine Use 2012, Pew Internet & American Life Project, Mar 9, 2012, http://pewinternet.org/Reports/2012/Search-Engine-Use-2012.aspx , accessed October 10,2012,